DOCUMENT RESUME

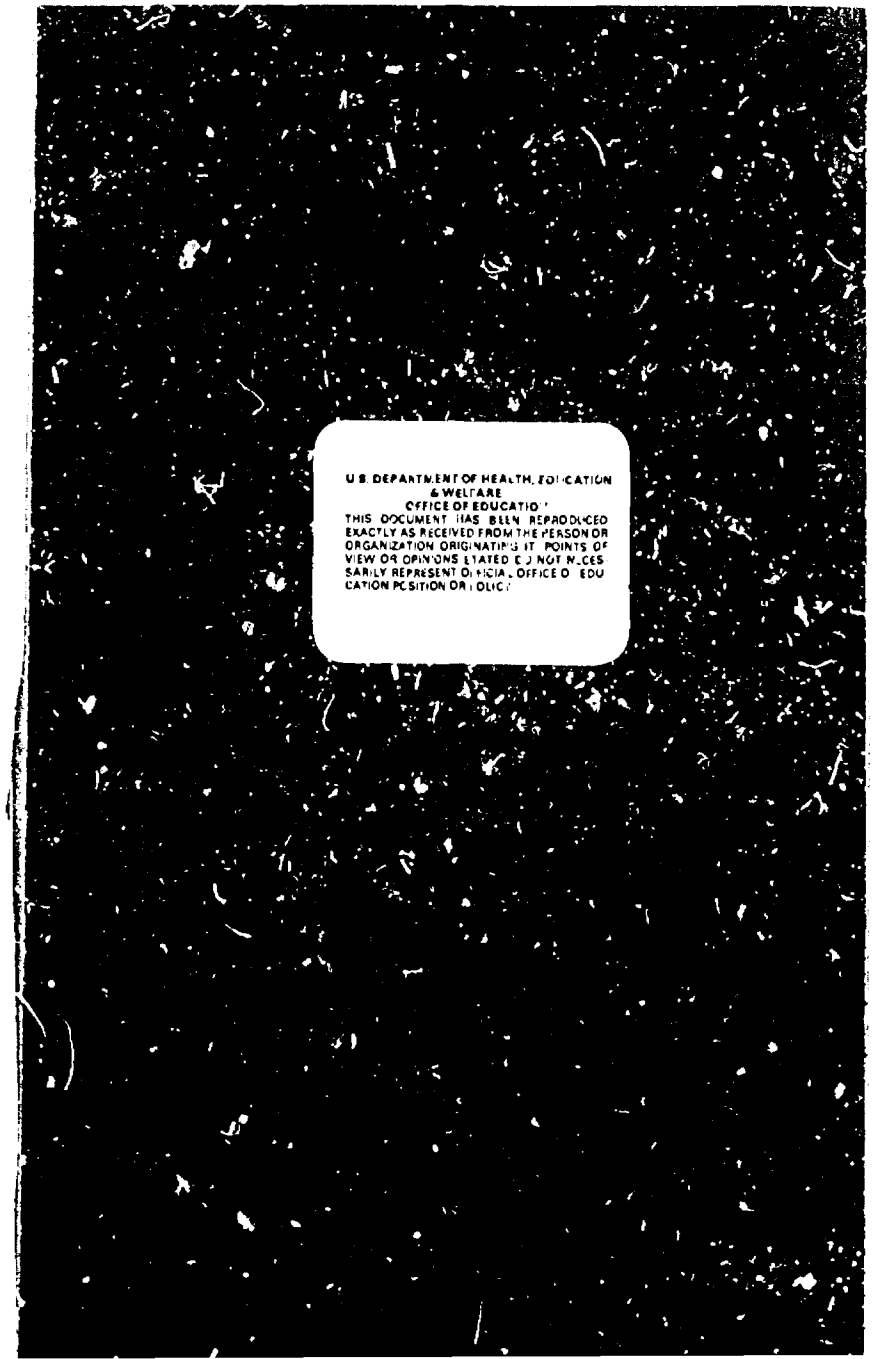
ED 051 284 TM 000 600

AUTHOR Bartos, Bruce B.
TITLE A Review of Instruments Developed to be Used in the Evaluation of the Adequacy of Reported Research.
INSTITUTION Phi Delta Kappa, Bloomington, Ind.
REPORT NO OP-No-2
PUB DATE Feb 69
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, California, February 1969

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Educational Research, *Evaluation Criteria, Evaluation Techniques, *Research Reviews (Publications), *Technical Reports, *Test Reviews

ABSTRACT

Three sources of information useful in evaluating the adequacy of reported research are discussed: articles, checklists, and rating scales. The chronological and genealogical relationships among some of these sources, their general approach, and some considerations for their use are indicated. A bibliography of 42 sources is provided. (DG)



A REVIEW OF INSTRUMENTS DEVELOPED TO BE USED IN THE EVALUATION OF THE ADEQUACY OF REPORTED RESEARCH

by

Bruce B. Bartos
Phi Delta Kappa & Indiana University

The three-fold purpose of this review will be to (1) provide a list of articles, checklists, and rating instruments, (2) show the chronological and genealogical relationships among some of the instruments, and (3) indicate the general approaches of these various types of instruments along with some crucial considerations for their use.

Purpose (1) has been satisfied by the provision of the 38 references on the appended list at the end of this paper. Of these, I have reproductions of 32. They have been arranged in chronological rather than alphabetical order.

Gephart, who is chairing this Symposium, compiled an extensive bibliography and instrument collection in the course of completing his doctoral dissertation entitled, "The Development of an Instrument for Evaluating Research Reports." The instrument therein, appropriately called the "Research Evaluation Instrument," was one of three minutely examined by Caroline S. Hodges, at the Bureau of Applied Social Research, in her Master's Thesis. Hodges located still more efforts in this direction. It was

these combined references that formed the nucleus for and gave impetus to my endeavors.

Chronologically, output in this area followed an interesting pattern from 1922 to 1967. In the 30 years from 1922-52 only six instruments were developed. This was followed by double that (or twelve) in the next 10 years, 1952-1962. This total of 18 instruments in 40 years can now be compared with 20 developed in the next five years alone. At this rate, it seems reasonable, and frightening, to estimate that by 1970 20 more instrument development efforts will have required the energy of educational researchers.

Genealogically, things are not quite so clear-cut. Less than half of these studies have bibliographic references included. Without these references a tracing of their antecedents becomes rather difficult. Of note, though, are the institutional influences of The Ohio State University and Columbia University, New York. The Ohio State University served as base for Clark, Guba, Kapfer, Cook, Gephart, Schneider, and Cady. Clark, Guba, and Smith have been mutually influential, even to the extent that all three are now at Indiana University. Gephart, after moving to the University of Wisconsin-Milwaukee, produced papers jointly with Ingle and Remstad. Similarly, Columbia University has known Bexler, Symonds, Nasatir, Sieber, Hodges, and Joel and Jean Davitz. These two groups account for 19 or 50% of the studies found.

When the documents are arranged by type, they divide into three categories: articles, checklists, and rating instruments. The operational definition of these categories hinges on their manner of indicating the guidelines by which research reports should be evaluated. The articles are expository in nature, employing declarative sentences, and imbeding the guidelines in the body of the text. Checklists are columns of questions, usually subsumed under criterial headings, and requiring only that one consider whether the question applies to some particular report. Instruments come equipped with multi-level, multi-faceted scales upon which one locates his answers to a string of questions or

statements. However, while the categories may seem to be neat, square boxes, the documents have round corners and bulging sides.

Typical of the article type of approach is Fox (12) who lists seven "criteria" including the purpose of the research; the research procedures; the research design; the limitations of the design; the analysis of the data; the conclusions; and the experience of the investigator. He closes his article with statements to the effect that while this is not an extensive list, "...if, however, the criteria help to make readers more constructively critical of reported findings of research, they will have served their purpose." Perdew (6) and Spence (19) follow a similar style. Perdew felt style of reporting should be a prime consideration, while Spence headed his list with a cautionary note about investigator's credentials.

It will be noted that earlier I put quotation marks around the word "criteria" with reference to Fox. This was because I feel almost none of the articles or instruments uncovered in my investigations completely satisfy the definition for a criterion. A criterion is a standard, like an inch in the measurement of length, having a zero baseline, and composed of finite increments. In lieu of the precise measurements of Physics, research report criteria should be arrived at through examples of consensually designated good and bad items.

Gephart (25) and Ingle and Gephart (32) are examples of a more adequate way to present criteria. The Gephart article discusses over thirty distinct criteria for methodological adequacy. Using those criteria, Ingle and Gephart do a pretty fair job of constructively critiquing a piece of research under the three headings: The Hypothesis, the Evidence (data), the Inference Pattern (logical structure). Hodges (30) also developed baselines for the comparison of her criteria.

Fully a third of the documents referred to herein are checklists. Most are a series of questions arranged in table form though a few are positive statements. They range from

Gibboney's (17) seven item gauge used to include or exclude research for a further review to Symonds' (9) 143 questions with which his Educational Psychology department reviewed dissertation proposals. It is misleading, however, to leave you believing that Symonds has that many distinct elements when actually they are grouped under 13 separate headings. The majority of the lists use between 5 and 15 main headings, and these headings tend to be the same ones Fox, Perdew, and Spence employ.

The most unique approach can be seen in the hybrid "checkstrument" (my word) developed by Smith (15). First, he assesses the *inadequacies*, not the attributes - probably in the belief that, as with whole cloth, it is easier to spot the flaws than it is to praise the completed product. Second, he supplies a five item code which ranges from the "inapplicability" of the inadequacy to "its presence is a serious flaw." Next, Smith gives examples as aids to answer most of his 52 (and numerous sub-) questions. And finally, he, like Gephart (22) and Nasatir (18, 29), provides an overall evaluation question.

Stephens (36), and to a lesser degree Smith (15), constructed a programed decision tree approach for their instruments. While Smith's is implicit, the explicit flow chart draw by Stephens clarifies and unifies his checklist's use. In making the point that just as the most competent researcher may occasionally do a poor piece of research, so might the "dufter, by good luck, come up with a useful product," Stephens emphasizes outcomes rather than procedures and methodologies. In this he differs from all his predecessors.

Although almost any instrument developed for the evaluation of *completed* research could likewise be used at the *planning* and *proposal* stages, only five of the authors take care to point this out. If consumers of research are going to be encouraged to use the various aids mentioned in this paper, they should be fully appraised of the context of their development and possible alternative uses. Symonds (9) perpared his guide for persons in educational psychology; Schneider and Cady (27) were concerned about music research; Kapfer - science education; and Suydam -

primary school arithmetic.

The last differentiating set of characteristics I will mention today is the type of research for which the instrument is intended. Johnson (10) recommends his rating instrument for both survey and experimental research, while Cook (21) limits his to experimental research only. Gephart, Ingle, and Remstad (33) analyzed comparative studies, and Perdew (6) felt that historical research should come under more methodical scrutiny.

I hope it has become apparent that the evaluation of research, both as proposals and final reports, has had a lengthy and varied history. The search for and development of adequate instruments must be continued. And interested members of AERA seem best suited to the task.

REFERENCES

1. Brooks, F. O. "Criteria of Educational Research," SCHOOL AND SOCIETY, 18:724-729; December 22, 1923

2. Bexler, H. H. "Checklists for Educational Research," Teacher's College, Columbia University, New York, p. 85-87, 1928

3. Wilson, G. M. "Research: Suggested Standards for Summarizing and Reporting Applied to Two Recent Summaries of Studies in Arithmetic," JOURNAL OF EDUCATIONAL RESEARCH 28:187-194; November, 1934

4. Scates, Douglas E. and Hoban, Charles F., Jr. "Critical Questions for the Evaluation of Research," JOURNAL OF EDUCATIONAL RESEARCH 31:241-254; December, 1937

5. Wolfle, Dael L. et al "Standards for Appraising Psychological Research," AMERICAN PSYCHOLOGIST 4:320-328; August, 1949

6. Perdew, Philip W. "Criteria of Research in Educational History," JOURNAL OF EDUCATIONAL RESEARCH 44:217-223; November, 1950

7. Record Form & Rating Scale for Evaluation of Graduate Research, American Institute for Research, 1955

8. Dvorak, Earl A. "General Guide to a Study of Research Reports," PEABODY JOURNAL OF EDUCATION 34:141-144; November, 1956

9. Symonds, Percival M. "A Research Checklist in Educational Psychology," JOURNAL OF EDUCATIONAL PSYCHOLOGY 47:100-109; February, 1956

10. Johnson, Garnville, B., Jr. "A Method for Evaluating Research Articles in Education," JOURNAL OF EDUCATIONAL RESEARCH 51:149-151; October, 1957

11. Van Dalen, Deobold B. "A Research Checklist in Education," EDUCATIONAL ADMINISTRATION & SUPERVISION 44:174-181; May, 1958

12. Fox, James H. "Criteria of Good Research," PHI DELTA KAPPAN 39:284-286; March, 1958

13. Farquhar, Wm. K. and Krumboltz, John D. "A checklist for Evaluating Experimental Research in Psychology and Education," JOURNAL OF EDUCATIONAL RESEARCH 52:353-354; May, 1959

14. Clark, D. L., Guba, E. G., Smith, G. R. "Functions & Definitions of Functions of a Research Proposal or Research Report," 1961

15. Smith, G. R. "Checklist of Inadequacies," November 17, 1961

16. Van Dalen, Deobold B. "Evaluation and Publication of Research" Chapt. 16 in UNDERSTANDING EDUCATIONAL RESEARCH, McGraw-Hill. 1962

17. Gibboney, Richard "Criteria for the Selection of Significant Educational Research Studies," (Prepared as material for guidance in developing a chapter in the REVIEW OF EDUCATIONAL RESEARCH, but not included and published) Robert B. Hayes and Emanuel Berger aided in this. August, 1963

18. Nasatir, E. David "David Nasatir's Detailed Instrument for Rating Educational Research Reports." Appendix I-A in MEASURING EDUCATIONAL RESEARCH QUALITY AND ITS CORRELATES by Caroline S. Hodges (Unpublished Master's Thesis, Columbia University, New York.

May 15, 1967.) Nasatir's work - 1963

19. Spence, Wm P. "Interpreting and Evaluating Research," Chapt. 4 of AMERICAN COUNCIL OF INDUSTRIAL ARTS TEACHER EDUCATION YEARBOOK 13:113-122; 1964

20. Kapfer, Philip G. "Criteria for Evaluating Research," 1964

21. Cook, Desmond L. "Checklist for Evaluating Design Aspects of Experimental Research" (There are two editions), 1964

22. Gephart, Wm J. "The Research Evaluation Instrument (REI)," 1964

23. Gephart, Wm J. "Development of an Instrument for Evaluating Educational Research Reports" (A Summary of CRP Project S-014), 1964

24. Gephart, Wm J. "REI Items that Discriminate Between "good" and "poor" Research in Counseling and Guidance," 1964

25. Gephart Wm J. "Criteria for Methodological Adequacy for Research on Educational Change," 1965

26. Schneider, E. and Cady, Henry L. "Evaluation & Synthesis of Research Studies Relating to Music Education," CRP E-016, Ohio State U., Columbus, Ohio, 1965, p. 36-43

27. Schneider, E. and Cady, Henry L. "Evaluation & Synthesis of Research Studies Relating to Music Education," CRP E-016, Ohio State U., Columbus, Ohio, 1965 (Exhibits 9 & 10: Abstractor's Summary and Abstractor's Guidebook, p. 631-642)

28. Sieber, Sam "Detailed Criteria of Evaluation: 1st and 2nd Revisions of Nasatir's Instrument" 1965, Appendix I-C of MEASURING EDUCATIONAL RESEARCH QUALITY

AND ITS CORRELATES by Caroline S. Hodges (Persell), Columbia U., New York, May, 1967

29. Nasatir, E. David "Nasatir's Global Rating Form" 1965, Appendix 1-F of MEASURING EDUCATIONAL RESEARCH QUALITY AND ITS CORRELATES by Caroline S. Hodges (Persell), Columbia, U., New York, 1967

30. Hodges, Caroline S. (Persell), "Caroline Hodges' Research Evaluation Instrument: Detailed Criteria Rating form (Revised version No. 2, July, 1966) Appendix 1-D of MEASURING EDUCATIONAL RESEARCH QUALITY AND ITS CORRELATES by Caroline S. Hodges, Columbia U., New York, May 15, 1967

31. Wandt, E. A CROSS-SECTION OF EDUCATIONAL RESEARCH, David McKay Publishers, 1966

32. Ingle, R. B. and Gephart, Wm J. "A Critique of a Research Report: Programed Instruction versus Usual Classroom Procedures in Teaching Boys to Read," AMERICAN EDUCATIONAL RESEARCH JOURNAL 3:49-53, 1966

33. Gephart, Wm J., Ingle, R. B., and Remstad, R. "A Framework for Evaluating Comparative Studies," In USOE Project 6-1388 Report: A Conference on Research in Music Education, Henry Cady, Ohio State U., Columbus, Ohio, May, 1967

34. Davitz, Joel R. and Lois, J. A GUIDE FOR EVALUATING RESEARCH PLANS IN PSYCHOLOGY AND EDUCATION, Teachers College Press, New York, 1967, 38 pp.

35. Suydam, Marilyn N. "An Instrument for Evaluating Experimental Educational Research Reports." Pennsylvania State Department of Public Instruction, 1967

36. Stephens, J. M. "Making Dependable Use of Published Research: A Proposed Checklist," THE JOURNAL OF

EDUCATIONAL RESEARCH 61:99-104; November, 1967

37. Chall, Jeanne S. "Schedules Used in Study," Appendix B in LEARNING TO READ: THE GREAT DEBATE, McGraw-Hill New York, 1967, 372 p.

38. Krahmer, Edward Questionnaire attached to Letter dated January 6, 1967, for a Research Project of the Bureau of Educational Research and Services, University of North Dakota, Grand Forks, 1967

39. Gephart, William J. "Profiling Educational Research." Occasional Paper No. 3 Phi Delta Kappa Research Service Center: Bloomington, Ind. January 1969

40. Gephart, William J. & Bruce B. Bartos, "Profiling Instructional Package." Occasional Paper No. 7, Phi Delta Kappa Research Service Center. Bloomington, Indiana, August 1969

41. Tripodi, Tony, Phillip Fellin, & Henry J. Meyer, THE ASSESSMENT OF SOCIAL RESEARCH. Itasca, Illinois, F. E. Peacock Publishers, Inc. 1969

42. Karlene H. Roberts "Understanding Research: Some Thoughts on Evaluating Completed Educational Projects." ERIC Clearinghouse on Educational Media and Technology, Stanford University, Stanford, California. July 1969